

Reporte de Investigación

Alex Medina

30/11/2022

Indice de contenidos

Introducción	1
Definición del problema	2
Obtención de datos	4

Introducción

El esplendor de la era digital han traído consigo un crecimiento exponencial de los repositorios del contenido multimedia digital. En 2022, la cantidad de usuarios de teléfonos inteligentes en el mundo actual es de 6648 millones, lo que se traduce en que el 83 % de la población mundial posee un teléfono inteligente, como consecuencia de esta distribución tan amplia y con la calidad cada vez mayor de las cámaras de los teléfonos inteligentes, la cantidad de fotografías tomadas en todo el mundo cada día se está disparando. En 2022, se toman 54.4 mil fotos por segundo y 1.7 billones por año.

Este desarrollo ha propiciado la creación de repositorios de imágenes bastante voluminosos de todos los campos semánticos imaginables. Estos conjuntos de datos son cada vez mas aprovechados por la industria publica y privada dentro del area de vision por computadora y aprendizaje automático para entrenar algoritmos de inteligencia artificial que que luego pueden ser aprovechados para lograr la identificación de patrones y clasificación sobre los conjuntos de images que cada vez aumentan más.

Definición del problema

Para el caso particular de esta investigación, el area de aplicación con la que se trabajará es el de prendas de vestir. Es posible resumir de manera más concreta el problema como:

Lograr la segmentación de prendas de vestir variadas por medio de la utilización de algoritmos de Vision por Computadora y *Deep Learning*. Esto implica la creación y optimización de modelos de clasificación que serán aplicados sobre un conjunto de imágenes que contienen una prenda de vestir específica.

Actualmente las soluciones para llevar a cabo la segmentación, ademas de la alternativa básica que consiste en utilizar a una persona para clasificar lo que puede observar en las imágenes, es la de la utilización de herramientas ya existentes similares de vision por computadora capaces de lograr dicha segmentación. Un ejemplo de estas herramientas es Google Cloud AutoML Vision, el cual es una herramienta de paga y en línea que puede ser utilizada incluso sin la necesidad de conocer sobre temas de desarrollo o programación para aprender como funciona.

Es posible representar el flujo del proceso de la solución de la siguiente manera:

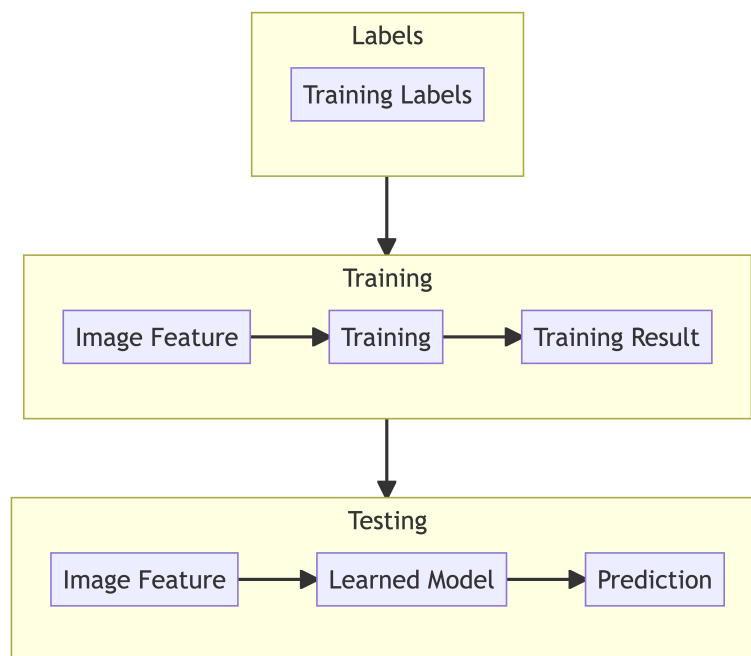


Figura 1: Flujo de trabajo del modelo.

El modelo será entrenado con un conjunto de imágenes de prendas de vestir variadas y etiquetas. Una vez que el modelo de entrenamiento sea lo suficientemente capaz de realizar

la identificación de las prendas en las imágenes, este mismo modelo será probado sobre un conjunto nuevo.

Para el entrenamiento de los modelos se utiliza un conjunto de datos que contienen etiquetas que identifican el tipo de prenda que está representada en la imagen a analizar, por se puede definir la investigación como un problema supervisado. Además es posible catalogarlo como un problema *online*, debido a que se pretende tener los datos completos a utilizar ya almacenados e ir mejorando la solución y los modelos con nuevos datos.

La manera en la que se estará midiendo el desempeño de los modelos es por medio de métricas básicas, se estará midiendo principalmente a través de el porcentaje de predicciones correctas que el modelo logre realizar sobre un conjunto de datos de prueba después de realizar el entrenamiento. Además también se pretende tomar en cuenta el tiempo de entrenamiento de los modelos ya que cada algoritmo puede tener un desempeño similar en cuanto a resultados pero diferente en cuanto a poder computacional requerido. Por lo tanto la selección de modelos estará basado en el equilibrio de estos dos parámetros.

Si se toma como base el trabajo que una persona podría realizar al intentar llevar a cabo esta segmentación de prendas de vestir, entonces resulta bastante simple medir el desempeño que los modelos logren, tanto por el porcentaje de predicciones correctas como por el tiempo en el que logre completar la clasificación.

Un humano con desarrollo cognitivo normal es capaz de realizar la segmentación e identificación de prendas de vestir si observa una imagen de ella con relativa facilidad. Se estima que el ojo de una persona experta en el tema llega a lograr un error de aproximadamente el 5%, esto al realizar el trabajo manual de segmentación e identificación de los objetos que se muestran sobre una imagen sobre un dataset de 1500 imágenes y 1000 clases, estos resultados son obtenidos del Large Scale Visual Recognition Challenge (ILSVRC) en el cual se ponen a prueba la experticia humana contra los modelos de vision por computadora. Por lo tanto al el objetivo es obtener al menos un 95% de predicciones correctas que representaría un desempeño similar al de una persona.

Una característica interesante de este tipo de soluciones es su escalabilidad y adaptabilidad para poder ser utilizado en problemas similares. La solución que se desarrollará debería ser posible de utilizarse con con problemas similares. Tanto la experiencia obtenida y las herramientas que se obtengan en este problema se pueden trasponer relativamente fácil sobre conjuros de datos similares. Esta es una ventaja ampliamente conocida en el area de vision por computadora, pues existen ya varios modelos e investigaciones que realizan un procedimiento similar pero aplicado sobre conjuros de datos diferentes o con contexto diferentes.

Lista las asunciones que has hecho hasta ahora.

- El conjunto de datos es usable.

Verifica si las asunciones se cumplen.

- ☒ El conjunto de datos es usable.

Obtención de datos

El conjunto de datos que será utilizado como para entrenar los modelos a probar fue obtenido desde [Kaggle](#), la cual es una plataforma para practicantes y profesionales del area de Ciencia de Datos donde los usuarios pueden publicar conjuntos de datos para realizar exploracion de datos, construccion de modelos, así como participar en competencias entre los mismo usuarios.

En este caso el repositorio que será utilizado en esta investigacion se puede descargar desde el siguiente [link](#).



Figura 2: Muestra del contenido del conjunto de datos

El conjunto de datos Pertenece a Alexey Grigorev y fue creado como un esfuerzo de contribución comunitaria, es decir que muchas personas se involucraron aportando sus propias imágenes. Se encuentra publicado bajo una licencia Creative Commons Zero (CC0), esto significa que cualquier individuo puede utilizar estos datos para cualquier fin, incluso comercial.

Más información sobre la obtención de los datos puede ser consultada desde el siguiente “[artículo](#)” escrito por el mismo usuario que publicó el conjunto de datos en Kaggle.

El tipo de datos son 10,000 imágenes de prendas de vestir en formato JPG

Dicho conjunto de datos tiene un tamaño de aproximadamente 7GB, el cual consta de un conjunto de 10000 imágenes catalogadas en 20 tipos diferentes de prendas de vestir, dividido en dos subconjuntos de 5000 imágenes cada uno, donde el primer conjunto son las imágenes en resolución original y el segundo en una version comprimida.

Los datos ya están en un formato manejable:

- Las imágenes se encuentran en formato JPG.
- Las etiquetas de las imágenes se encuentran en un archivo CSV.

No existe información sensible, por lo que este paso ha sido excluido.

Para obtener los datos de manera automática se hace uso de la API (Application Program Interface) que proporciona el propio Kaggle. Lo único necesario es generar un *token* o firma cifrada que permite una autenticación dentro del sitio y así poder descargar el repositorio.